REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-10-2012	Final Report	September 2008-August 2011

4. TITLE AND SUBTITLE		5a. CONTRACT NUMBER
Optimization Techniques for Clustering, Connectivity, and Flow Problems in Complex Networks		FA9550-08-1-0483
		5b. GRANT NUMBER
		FA9550-08-1-0483
		5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Sergiy Butenko	
Vladimir Boginski	**5e. TASK NUMBER**
Oleg Prokopyev	
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Dept. of Industrial and Systems Engineering, Texas A&M University, 4037 Emerging Technologies Build., College Station, TX 77843-3131, Phone: 979-458-2333, Fax: 979-458-4299, Email: butenko@tamu.edu	N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
875 North Randolph St.	AFOSR
Suite 325, Rm.3112	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Arlington, VA, 22203-1768	AFRL-OSR-VA-TR-2012-0882

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project develops network-based optimization methods for solving problems arising in complex system analysis and wireless networking applications. The study of complex systems is of utmost importance for a number of diverse areas of science, engineering and society, including biochemistry, social sciences, epidemiology, transportation, and telecommunications. The project makes contributions to the state of the art of network-based techniques for data mining of complex systems and virtual backbone-based routing in wireless ad hoc networks. The research in this project focused around the following four major thrusts: (I) Theoretical analysis of new models of clusters in networks; (II) Investigating new approaches to virtual backbone-based routing in wireless networks; (III) Establishing techniques for theoretical analysis of heuristics for inapproximable problems and designing new metaheuristic approaches for the problems of interest; (IV) Developing new models and algorithms for robust optimization and decision making in complex networks under uncertainty.

15. SUBJECT TERMS

Complex Networks, Optimization, Risk and Uncertainty

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Sergiy Butenko
				5	**19b. TELEPHONE NUMBER** *(Include area code)*
					(979) 458-2333

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Optimization Techniques for Clustering, Connectivity, and Flow Problems in Complex Networks

PI: Sergiy Butenko

Dept. of Industrial and Systems Engineering, Texas A&M University, 4037 Emerging Technologies Build.
College Station, TX 77843-3131, Phone: 979-458-2333, Fax: 979-458-4299, Email: butenko@tamu.edu

Co-PI: Vladimir Boginski

UF-REEF and Industrial and Systems Engineering, University of Florida, 1350 N Poquito Road
Shalimar, FL 32579, Phone: 850-833-9355 ext 240, Fax: 850-833-9366, Email: boginski@reef.ufl.edu

Co-PI: Oleg Prokopyev

Dept. of Industrial Engineering, University of Pittsburgh, 1048 Benedum Hall
Pittsburgh, PA 15261, Phone: 412-624-9833, Fax: 412-624-9831, E-mail: prokopyev@engr.pitt.edu

Final Project Report

09/01/2008 – 08/31/2011

Abstract

This project develops network-based optimization methods for solving problems arising in complex system analysis and wireless networking applications. The study of complex systems is of utmost importance for a number of diverse areas of science, engineering and society, including biochemistry, social sciences, epidemiology, transportation, and telecommunications. The project makes contributions to the state of the art of network-based techniques for data mining of complex systems and virtual backbone-based routing in wireless ad hoc networks. The research in this project focused around the following four major thrusts: (I) Theoretical analysis of new models of clusters in networks; (II) Investigating new approaches to virtual backbone-based routing in wireless networks; (III) Establishing techniques for theoretical analysis of heuristics for inapproximable problems and designing new metaheuristic approaches for the problems of interest; (IV) Developing new models and algorithms for robust optimization and decision making in complex networks under uncertainty. More specific contributions resulting from this project include: (I) Introducing and investigating a robust version of the concept of k-club in networks, which embeds the requirement of predefined vertex connectivity into low-diameter structures; studying the maximum k-club and maximum quasi-clique problems from mathematical programming perspective. (II) Developing new approaches to forming a virtual backbone in wireless sensor networks by introducing and utilizing a new concept of a bottleneck connected dominating set, which help to address the energy-scarcity considerations; distributed algorithms for approximating a minimum bottleneck connected dominating set are proposed. (III) Providing a framework for theoretical analysis of heuristics for hard discrete optimization and for analysis of performance of algorithm portfolios; introducing a metaheuristic framework of variable objective search that can be used for discrete optimization problems with different mathematical programming formulations, including continuous formulations. (IV) Suggesting LP-based solution methods for network flow problems subject to multiple uncertain arc failures, as well as for rehabilitation and maintenance decisions related to transportation infrastructure.

1 Summary of Research Contributions

The research contributions resulting from the project in its four major thrusts are summarized in the following subsections. Namely, we list abstracts/brief summaries of the papers supported by this project that have been finalized or are close to completion. Publication [11] is not directly related to the project, however, it resulted from the collaboration facilitated by this grant.

1.1 Theoretical analysis of new models of clusters in networks

The results of our work dealing with theoretical, as well as computational analysis of new models of clusters in networks are reported in [1, 2].

In [1], we deal with variations of the maximum k-club problem in networks. Network robustness issues are crucial in a variety of applications. In many situations, one of the key robustness requirements is the connectivity between each pair of nodes through a path that is *short enough*, which makes a network cluster more robust with respect to potential network component disruptions. A *k-club*, which by definition is a subgraph of a diameter of at most k, is a structure that addresses this requirement (assuming that k is small enough with respect to the size of the original network). We develop new compact linear 0-1 programming formulations for finding maximum k-clubs that substantially reduce the number of entities compared to the known formulations ($O(kn^2)$ instead of $O(n^k)$), which is important in the general case of $k > 2$. Moreover, we introduce a new related concept referred to as an *R-robust k-club* (or, k_R-club), which naturally arises from the developed k-club formulations and extends the standard definition of a k-club by explicitly requiring that there must be at least R distinct paths of length at most k between all pairs of nodes. A compact formulation for the maximum R-robust k-club problem is also developed, and error and attack tolerance properties of the important special case of R-robust 2-clubs are investigated. Computational results are presented for power-law and uniform random graph instances.

In [2], we study the maximum quasi-clique problem, which defines a cluster based on edge density. Given a simple undirected graph $G = (V, E)$ and a constant $\gamma \in (0, 1)$, a subset of vertices is called a γ-quasi-clique or, simply, a γ-clique if it induces a subgraph with the edge density of at least γ. The maximum γ-clique problem consists in finding a γ-clique of largest cardinality in the graph. Despite numerous practical applications, this problem has not been rigorously studied from mathematical perspective, and no exact solution methods have been proposed in the literature. This paper, for the first time, establishes some fundamental properties of the maximum γ-clique problem, including the NP-completeness of its decision version for any fixed γ satisfying $0 < \gamma < 1$, the weak heredity property, and analytical upper bounds on the size of a maximum γ-clique. Moreover, mathematical programming formulations of the problem are proposed and results of preliminary numerical experiments using a state-of-the-art optimization solver to find exact solutions are presented.

1.2 New approaches to virtual backbone-based routing

New approaches to virtual backbone-based routing utilizing the concept of a bottleneck dominating set have been proposed and studied in [3, 4].

Given a simple undirected graph, the minimum connected dominating set problem is to find a minimum cardinality subset of vertices D inducing a connected subgraph such that each vertex outside D has at least one neighbor in D. Approximations of minimum connected dominating sets are often used to represent a *virtual routing backbone* in wireless networks. The paper [3] first proposes a constant-ratio approximation algorithm for the minimum connected dominating

set problem in unit ball graphs and then introduces and studies the edge-weighted bottleneck connected dominating set problem, which seeks a minimum edge weight in the graph such that the corresponding bottleneck subgraph has a connected dominating set of size k. In wireless network applications this problem can be used to determine an optimal transmission range for a network with a predefined size of the virtual backbone. We show that the problem is hard to approximate within a factor better than 2 in graphs whose edge weights satisfy the triangle inequality and provide a 3-approximation algorithm for such graphs. We also show that for fixed k the problem is polynomially solvable in unit disk and unit ball graphs.

In [4], we further develop the idea of using bottleneck connected dominating sets as a more energy-aware alternative to the minimum connected dominating sets in routing protocols for wireless sensor networks. The traditional approach is to assume that the transmission range of each node is given and to minimize the number of nodes in the connecting dominating set representing the virtual backbone. Using the concept of k-bottleneck connected dominating set (k-BCDS), which, given a positive integer k, minimizes the transmission range of the nodes that ensures a CDS of size k exists in the network, allows to vary the transmission range of nodes to facilitate a routing protocol based on a virtual backbone of a desirable size. This paper provides a 6-approximate *distributed* algorithm for the k-BCDS problem. The results of empirical evaluation of the proposed algorithm are also included.

1.3   Theoretical analysis of heuristics and designing new metaheuristic approaches

Papers [5–7] describe our contribution to theoretical analysis and design of heuristic approaches to hard discrete optimization problems.

In [5], we examine the bound on effectiveness of using algorithm portfolios and restart strategies. Randomized heuristic algorithms constitute an effective approach for solving a wide range of computationally challenging problems including various large-scale combinatorial optimization problems. These methods can often be accelerated by restarting the randomized algorithm each time its running time exceeds some predefined threshold. Given a set of available algorithms for a particular problem, one can either choose the best performing algorithm and run multiple copies of it in parallel with different random seeds (single algorithm portfolio), or choose some subset of algorithms to run in parallel (mixed algorithm portfolio). It has been previously shown in the literature that the latter approach may provide better results. We investigated the extent of such improvement. We provided a theoretical bound on the maximum ratio of expected running times corresponding to each of these parallel approaches.

In [6], we propose an approach to justify the usage of certain greedy heuristics that extends the common definition of the "best approximation algorithm" to the problems for which a constant-ratio approximation algorithm is unlikely to be found. Namely, given a problem $\mathcal{P}$ and a heuristic algorithm $\mathcal{A}$ for this problem, we can claim that this heuristic is the "best" for the given problem if finding a solution better than that output by $\mathcal{A}$ (whenever one exists) is NP-hard. To prove such a result it suffices to show that it is NP-hard to recognize whether there is a gap between the optimal objective function value and the value of the solution output by $\mathcal{A}$. We identify provably best (according to this definition) heuristics for several important problems in networks, including the maximum clique, maximum independent set, maximum k-plex, and maximum k-club problems.

In [7], we introduce the variable objective search framework for combinatorial optimization. The method utilizes different objective functions used in alternative mathematical programming formulations of the same combinatorial optimization problem in an attempt to improve the solutions obtained using each of these formulations individually. The proposed technique is illustrated using alternative quadratic unconstrained binary formulations of the classical maximum independent set

3

problem in graphs.

Article [8] investigates the local maxima properties of a box-constrained quadratic optimization formulation of the maximum independent set problem in graphs. Theoretical results characterizing binary local maxima in terms of certain induced subgraphs of the given graph are developed. We also consider relations between continuous local maxima of the quadratic formulation and binary local maxima in the Hamming distance-1 and distance-2 neighborhoods. These results are then used to develop an efficient local search algorithm that provides considerable speed-up over a typical local search algorithm for the binary Hamming distance-2 neighborhood.

1.4   Optimization and decision making in complex networks under uncertainty

In [10], we propose LP-based solution methods for network flow problems subject to multiple uncertain arc failures, which allow finding robust optimal solutions in polynomial time under certain conditions. We justify this fact by proving that for the considered class of problems under uncertainty with linear loss functions, the number of entities in the corresponding LP formulations is polynomial with respect to the number of arcs in the network. The proposed formulation is efficient for sparse networks, as well as for time-critical networked systems, where quick and robust decisions play a crucial role.

In [9], a method for determining optimal risk-based maintenance and rehabilitation (M&R) policies for transportation infrastructure is presented. The proposed policies guarantee a certain performance level across the network under a predefined level of risk. The long-term model is formulated in the Markov Decision Process (MDP) framework with risk-averse actions and transitional probabilities describing the uncertainty in the deterioration process. The well known Conditional Value at Risk (CVaR) is used as the measure of risk. The steady-state risk-averse M&R policies are modeled assuming no budget restriction. To address the short-term resource allocation problem, two linear programming models are presented to generate network-level polices with different objectives. Comparison of the results of the two models suggests that distributing resources among facilities to minimize the sum of CVaR will result in a high variance among the CVaR level of individual facilities, while distributing resources to minimize the largest CVaR of the facilities will result in a slightly higher total sum of CVaR and considerably smaller variance.

References

[1] A. Veremyev and V. Boginski. Identifying large robust network clusters via new compact formulations of maximum k-club problems. *European Journal of Operations Research*, 218(2): 316–326, 2012.

[2] J. Pattillo, A. Veremyev, S. Butenko, V. Boginski. On the maximum quasi-clique problem. Submitted to *Discrete Applied Mathematics*, 2011.

[3] S. Butenko, S. Kahruman-Anderoglu, and O. Ursulenko. On connected domination in unit ball graphs. *Optimization Letters*, 5:195–205, 2011.

[4] A. Verma and S. Butenko. A distributed approximation algorithm for the bottleneck connected dominating set problem. *Optimization Letters*, 2011. To appear. DOI 10.1007/s11590-011-0314-2

[5] O.V. Shylo, O.A. Prokopyev, and J. Rajgopal. On algorithm portfolios and restart strategies. *Operations Research Letters*, 39(1): 49–52, 2011.

[6] S. Butenko, S. Kahruman-Anderoglu, and O. Prokopyev. On provably best heuristics for hard discrete optimization. *Working Paper*, 2012.

[7] S. Butenko, O. Yezerska, and B. Balasundaram. Variable objective search. *Journal of Heuristics*, 2011. To appear. DOI 10.1007/s10732-011-9174-2.

[8] F. Mahdavi Pajouh, B. Balasundaram, and O. Prokopyev. On characterization of maximal independent sets via quadratic optimization. *Journal of Heuristics*, 2011. To appear. DOI 10.1007/s10732-011-9171-5.

[9] S. R. Seyedshohadaie, I. Damnjanovic, and S. Butenko. Risk-based maintenance and rehabilitation decisions for transportation infrastructure networks. *Transportation Research A*, 44: 236–248, 2010.

[10] V. Boginski, C.W. Commander, and T. Turko. Polynomial-time Identification of Robust Network Flows under Uncertain Arc Failures. *Optimization Letters*, 3(3):461-473, 2009.

[11] O. Prokopyev, S. Butenko, and A. Trapp. Checking solvability of systems of interval linear equations and inequalities via mixed integer programming. *European Journal of Operational Research*, 199:117–121, 2009.